DOCUMENT RESUME

ED 103 557 UD 014 932

AUTHOR Franklin, Anderson J.
TITLE The Testing Dilemma for Minorities.
PUB DATE Oct 74
NOTE 15p.; Paper presented at the Public Hearings on Statewide Testing and Evaluation, State of New York (Albany, N.Y., October 1974)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Achievement Tests; Aptitude Tests; *Educational Policy; *Educational Testing; Government Role; *Minority Group Children; Public Policy; Racial Discrimination; Social Discrimination; *State Government; State Programs; *Test Bias; Testing Problems; Testing Programs; Test Validity

ABSTRACT

The document states that certain steps need to be taken immediately for rectifying and containing the injustices of testing. Until such time that the State can demonstrate unequivocally that their statewide testing and evaluation program is fair to all groups, and that every student has had an equal exposure to quality school environments before evaluation then there should be a moratorium on testing. The State should establish a task force for the development of an Office of Consumer Affairs in Testing and Student Evaluation. The State should establish a Research and Development Office which will have the latitude to study empirical questions of teacher and pupil performance. It is most important that evaluative agencies recognize that tests and their ensuing social judgments are instrumen s of racism by virtue of minority exclusion in all phases of test utilizations. Moreover since minorities have limited access to the opportunity (mainstream) structures of this society, much less policy making positions, it is obvious that decisions on criterion variables (job or education) have negligible minority inputs. Since racism has been an integral characteristic of the power brokers in this country, and the testing industry caters to the power brokers, there is no reason to assume that testing has the best interests of minorities at heart. (Author/JM)

ED103557

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BEST COPY AVAILABLE

The Testing Dilemma for Minorities

Anderson J. Franklin

Teachers College
Columbia University

UD 014 932

Paper presented at the public hearing on Statewide Testing and Evaluation, State of New York, Executive Chamber, Office of Education Performance Review, State Capitol, Albany, New York, October, 1974.

BEST COPY AVAILABLE

Introduction

In trying to prepare a statement on statewide testing it again brought home the indisputable fact that this evaluative process has enormous shortcomings and abundant complexities. However we continue to pursue this assessment technique with a relentlessness that has to bring our professional judgments into serious question. The debate over testing, its biases, its fairness, its consequences, continues incessantly. Like that rat on the activity wheel our learned discussions on the test issue is getting us nowhere fast. The only vivid truth about our testing practices is that individuals are having their lives charted for better or for worse by assessment techniques which are at best still primitive. Perhaps our serious error in judgment is the lack of control we exert over the utilization of these instruments. There is no effective monitoring of how these tests are employed. It is ludicrous to assume that admonitions written in test manuals will be strictly adhered to by the user. The fact that test limitations are a major psychometric issue is negligible for most agencies employing tests. The test user is seeking an expedient method of evaluation and selection. Our profession has provided that solution in the form of innumerable standardized instruments. But in supplying that demand we have neglected to impress upon the users the deficiencies of the tool placed in their hands. In effect the misuse and abuse of tests becomes the liability of the test professionals by default and abdication of responsible guidance in its use. Consequently we end up in perennial meetings like this one today trying to resolve a dilemma that seems insoluble by the very persistence of the issue throughout the history of testing. I attribute part of the problem to the inescapable social issues dominating the use of tests. Whenever a device is used to determine the status of individuals, discrimination occurs that carries just as much social implications for me as it does technical.

In my opinion the responsibility of this panel in considering the situation of statewide testing encompasses the review of both the technical and social implications of pupil evaluation. The Regent examinations, the Pupil Evaluation Program and all other statewide tests are subject to technical and social scrutiny. Parenthetically I include the growing legal challenges confronting the test industry within the social domain. We cannot overlook the stigma associated with poor performance on tests. The consequences of negative labeling to the personal development of the individual has been adequately articulated in the book Stigma by Goffman (1963). No responsible State authority would be party to such a process. Given the tests employed by the State plus the testing dilemma for minorities I cannot in good conscience absolve the State from participating in the unfair classification of minority students with their Statewide Testing program. For the protection of student welfare this is an area in need of careful scrutiny. It is one that I and other colleagues will begin to examine closely. But just as our efforts develop in this domain I would hope that the substance of my statement here today will begin to move the State into more careful consideration of the consequences of its Statewide Testing program to the development of the individual, not the economics of education or the whims of legislative and political demands. Consequently, I would like to put several questions before this panel which if solutions are seriously sought in response to them we may achieve movement towards a better educational system.

My questions are as follows:

1. In the current State testing program, what is the ultimate assessment purpose for every instrument used? Is it for expediency or education?

2. Given the tests employed by the State, what is the actual utility of them? Do they provide sufficient information on the school's role in student preparation and intellectual development? Do they contribute to educational planning or primarily classification in a meritocracy?

3. What is the range and adequacy of the criterion variables in evaluating student and school performance. Similarly to what extent has evaluation procedures been limited to quantitative assessment methods and criteria in contradistinction to qualitative variables, (e.g., motivation - teachers' and students')?

4. What is the interaction between school performance and student performance? Are there comparable measures of competency in the school as there are for the student? If not, then what is the value of judgments on the intellectual development of the pupil?

5. If the standardization process in the development of tests (particularly item selection) has not had sufficient minority group representation and professional input then on whose standard is evaluation being performed? Parenthetically, how is the differential development of specific groups as defined by sex, age, race, geographical location, social class, etc., considered in pupil evaluation? Moreover what is the State's responsibility in the fate of many children classified as educably retarded or slow learners because of culturally biased and unfair tests as well as inequity in the quality of school experiences?

6. Within the total scope of the statewide evaluative system to what extent is the ecology of learning and development for specific groups considered?

7. What is the status of fundamental research on education and assessment in the State Education Department? Furthermore, what has the Department contributed to our understanding of the learning process in students?

If we equivocate on providing substantial answers to these questions or if the answers are inadequate then what human right do we possess to pass judgment on and determine the lives of our children from flimsy pupil evaluation techniques?

Recommendations

In my opinion certain steps need to be taken immediately for rectifying and containing the injustices of testing. The following is a partial list of recommendations to improve Statewide responsibility in the area.

1. Until such time that the State can demonstrate unequivocally that their statewide testing and evaluation program is fair to all groups, and that every student has had an equal exposure to quality school environments before evaluation then I reaffirm the resolution of the Association of Black Psychologists and recently the NAACP and the National Education Association (see Journal of School Psychology, 1973)

which declares a moratorium on testing. This should not be construed that the search for alternative means of assessment be abandoned, but that the current testing practices be halted.

2. The State should establish a task force for the development of an Office of Consumer Affairs in Testing and Student Evaluation. The responsibilities of this Office would include the following:

   a. consumer advocacy regarding the use and misuse of tests as well as advising parents of their legal rights in the testing of their children.

   b. advocating for the adoption of a "Truth-in-Testing" law in the State Legislature.

   c. a test review board to scrutinize and systematically monitor test utility, development, policies and practices of all agencies employing assessment procedures.

   d. the development of informational advisory centers plus layman documents on testing for parents of school children.

   e. the development of comprehensive Statewide Standards on Testing which reflect the interests of minority groups on the testing issue.

3. The State should establish a Research and Development Office which will have the latitude to study empirical questions of teacher and pupil performance in contrast to an Office which functions as a statistics mill for legislative accountability.

As further food for thought the remainder of my comments on the testing dilemma for minorities come from the working supposition that the systematic exclusion of minorities at all phases of test development to utilization presents a major concern in the evaluation of minority intellectual development and subsequent achievements in life.

I. Statement of the Issues

Tests constitute a major dilemma for minorities. (See Williams, 1971; Jones, 1972; Flaugher, 1973) They often have performances which fall below the norm. Because of this performance many of their lives become subjected to below-standard opportunities. In schools they are tracted and in jobs they are screened out. The relegation of an individual to a less than opportunist position is to deny his human rights. However daily, momentous decisions of this kind are made for minority populations based on the frailty of standardized instruments. The reliance on test information in our society - and throughout the world - has achieved the status of an institutionalized practice. The demands of a highly industrialized and technocratic society requires expediency and this is what tests offer - expediency. But what price is paid for the rush toward evaluative decisions of this nature? It is usually the sacrificing of individuality plus the consideration of the human capacity for resiliency, adaptation and range of capabilities. I remember distinctly in a class with Leona Tyler, a distinguished Counseling and Measurement Psychologist and former President of the American Psychological Association, in which she postulated a theory of possibilities to characterize human potential and risks in development. The essential point of this theory was that an individual possesses a range of potential abilities. Its development and ultimate manifestation was in part determined by access to the proper experiences of life. Maximum development of a potential was contingent upon the nature of specific exposures experienced by the individual. The significance of this theory for a discussion of the testing issue is the social neglect of the individual's range of possibilities to perform by the specificity of test demands and the narrow-mindedness which ensues from test interpretations and standards.

This inevitably leads to the issue of tests abuse. For if tests are designed to differentiate then someone must fall in the relative position of lower status. For minorities the lower status tends to be the perennial position. The question is why and what of the consequences. The arguments are abundant but they continually return to the issue of culturally biased tests. One can spend an eternity debating the fine points of this position. As muddled as the issue is, there is some legitimacy to the argument that the diversity of cultural backgrounds in a sample population is given little consideration in test development. This is most reflected in the content of tests, its instructions and procedures of administration. Likewise, statistical procedures cannot be the principal accounting tool in the normalization process. I feel the major point to be understood in the "culture bias issues" is the fact that we know very little about how cultural ethos and ethnocentricism define the ecology of learning and development for groups of people. It is this concern which lends credence to the claim of cultural bias in tests.

Cultural fairness of tests is another rationale to explain the testing dilemma for minorities (Flaugher, 1973). Eells, Davis, Havighurst, et. al. (1950) attempted to examine cultural factors in test item responses and one of their conclusions is that our focus should be on how "fair" a test is to given populations. Such a focus does not remove the injustices tests heave upon the less fortunate minorities but it begins to shift our sights from the cultural validity of the test content to the predictive validity of an instrument. With greater representation of diverse populations in the standardization process and improved specification of criterion variables, prediction is refined. But this does not prevent the stigma of categorization from poor test performance. Although our tests may become more "fair" through sample representation, the sorting of group performances for prediction does not resolve the testing dilemma. Value judgments still dominate the determination of criterion variables.

It is important to remember the purpose of tests in this technocratic society -- i.e. screening. This not only includes the tests as screening tools but the perceived utility of the instrument by the test users. It is at this juncture that the establishment of cutt-off scores becomes of paramount importance. If a selective process were not involved, the purpose of testing would be negligible. However the primary role of testing is to sort into the "haves and the have-nots." Testing is a very elitist process in practical objectives. We take the "cream off the top" in order to fulfill our institutional requirements. The chance of having failures after sorting is reduced accordingly. For an instrument to achieve such a selective capacity there are a number of factors involved in its construction, administration and utilization. Within these three areas is the heart of the testing dilemma for minorities. To what extent are minority interests considered during these phases -- very often little. The purpose for a test is the primary instigator in its construction. It is at this point that the foundation for the development of a tool is articulated between the agency or group in need and the test developers. Within this sanctum the practical utility is delineated as well as measurement objectives. Needless to say that the adage "a test is no better than its developers" fits most appropriately at this point. On countless occasions when minority group bitterness over testing was at its ostensible zenith a common complaint was the exclusion of minority input during test development or even revisions. Consequently the dilemma of tests for minorities can be considered as beginning here.

But the dilemma becomes more confounded with the administration of tests. There are serious problems in this area also. The way a test administrator presents the task can be a formidable encounter. Testing is already shrouded with the stigma of sorting. The

individual knows performance weights heavily in the direction his future life will take. Apprehension is frequently the psychological state that individuals bring to the testing situation. It is a known fact that test anxiety can have a deleterious effect on performance (Sarason, 1960). Very often, however, test administrators in their presumed desire to conform to procedures disregard the elements of apprehension which frequently loom in the testing room. The atmosphere established by the administrator is too often impersonal. The measurement legitimacy of this posture is not so much questioned as is the utility of this affectation for engendering maximum performance from test takers. Within our every-day learning environments, both formal and informal, where performance for survival really counts the demands for production are not so artificial. Consequently, the testing environment turns off production rather than enhancing it. I do not mean to construe that this phenomena is limited to minority populations but the experience does have its peculiar reception by minority persons. However another aspect of test administration has a real differential impact on minority populations and that is the delivery of instructions by the test administrator. This is most poignantly represented by a group of students I work with in a high school equivalency program. Teachers in this program were distressed at the results of student scores on the equivalency examination. The number of failures was in marked contrast to what performance on work book tasks would have predicted. Pressed to determine the cause of such test performance, they eventually learned from careful interviewing and backtracking that students did not understand the test administrator adequately. The test administrator's affect and diction during presentation of instructions was inhibitive. I have no fears of reinforcing racist stereotypes when I say that affect and diction are an integral part of black people's communication processes as well as those of many other

minorities of color -- be they Hispanic or Native American. Given this fact there is a serious discontinuity between the formal and sterile test situation and that of the every-day communication arena. Some may say this is a debatable issue or irrelevant to the situational requirements; but the main point to consider is that test administration objectives may not be conducive for maximizing performance (Hertzig, Birch, and Thomas,1969).

One other issue within the domain of test administration is the language of instructions, and the permissibility to deviate from them to facilitate comprehension. Much to the surprise of my teacher friends in the equivalency program their flexbility in providing numerous interpretations of instructions in the work manual until student comprehension was indisputable contributed greatly to the successful performance. In this instance the language as well as style of presenting test instructions often becomes an obstacle to minority students, and thus the testing dilemma is compounded.

Utilization of tests is another area of central concern in the testing dilemma from minorities. This incorporates not only test interpretation and the ensuing consequences but also the process of test selection. The tests we select almost mandate its utilization objectives. This is, in part, due to the specific scope and assessment purpose of the instrument. Invariably the tests we select are for predictive purposes. Our interests are in determining the probability of an individual's success within our institution. Consequently, a dependency on the norms emerges for the judgment process (Cleary, 1968). The crucial issue in normative data is its representation of the population, its predictive value and the line of demarcation for cut-off scores. Thorndike (1971), Darlington (1971), Cole (1972) and other measurement specialists discussed the problems associated with cut-off scores and culture fairness of tests. The inclusion and exclusion of individuals by the cut-off score does not consider adequately the

slope of the regression curve for individual groups in relationship to the test and criterion variables. Consequently absolute dependency on the norms for selection decisions is less efficacious than an inclusion of a number of criteria, both quantitative and qualitative. There is still considerable over and under prediction from our assessment instruments. This is, in part, due to the fact that the success of individuals below the cut-off scores is still probabilistic. These are important technical issues that test users should be sensitive to but often unaware of when considering the utility of a selection instrument.

A growing complaint among the many community and professional groups is the incongruity between the task of tests and on-the-job task requirements, i.e. the specification of the criterion variable. In many agencies the tests employed have little relationship to the job. Personnel and admission offices have operated on the premise of the correlation between standardized instruments and the probability of job or school success. It is not often considered that poor performers on the test tasks may have the requisite capabilities for on-the-job performance. This, however, would require other assessment techniques. Since time and money are factors in developing any diversified and unique or specific assessment approach, it is convenient to stick with instruments with correlational value. The utility of this procedure may be expedient but it does raise the issue of fairness. Many potentially successful persons are being excluded by this technique. During the days of the guild system there was merit in the apprenticeship program because it at least allowed the individual to succeed or fail at the job he must master. To a degree the open admissions policy in some state college systems is offering the same kind of opportunity. It may reflect a sink or swim philosophy but at least the chance to plunge into the stream of activity is offered.

Assumptions in Testing

The validity of a test is perhaps one of the most heatedly discussed issues throughout lay and professional circles. It certainly is not an issue that can be easily resolved. On the other hand those of us who have been exposed to measurement courses have frequently been informed of certain objectives to obtain in test development regarding validity. A test is considered valid if it meets one or more criteria. Several kinds of validity concepts must be considered in the process of test development -- i.e., construct validity, content validity, predictive validity, face validity, and concurrent validity (Cronbach, 1970).

It is the content validity area which forms the substance of the culture bias issue. The cultural origin of task content does effect cognitive performance (Franklin, 1974). For the resisters of testing however, the major validity argument hinges on the exclusion of capable persons by tests regardless of the technical integrity of the instrument. Since opportunity selection has greater social than technical implications the latter argument of test resisters has considerable credence.

Conclusion

In just this brief review of some issues in the testing dilemma for minorities it is apparent that the problem is immense. It is greatly obfuscated by the social concerns of the lay community and the questioned professional integrity of measurement specialists. In spite of this conflict it is most important that evaluative agencies recognize that tests and their ensuing social judgments are instruments of racism by virtue of minority exclusion in all phases of test utilization. Moreover since minorities have limited access to the opportunity (mainstream) structures of this society much less policy making positions it is obvious that decisions on criterion variables (job or education) have negligible minority input. So our lives

are entrusted to the "goodwill" of the establishers of the criteria. There is nothing in the history of this country which makes me believe that benevolence has won over racism in the interests of minorities. Consequently, since racism has been an integral characteristic of the power brokers in this country, and the testing industry caters to the power brokers there is no reason to assume that testing has the best interests of minorities at heart.

References

Bosma, B. "The NEA Testing Moratorium." Journal of School Psychology, (Special Issue), 11, 4, 1973.

Cleary, T.A. "Test Bias: Prediction of Grades of Negro and White Students in Integrated Colleges." Journal of Educational Measurement, 1968, 5, 115-124.

Cole, N.S. "Bias in Selection." ACT Research Report No. 51, Iowa City, Iowa: American College Testing Program, 1972.

Cronbach, L.J. Essentials of Psychological Testing, (3rd ed.), New York: Harper & Row Publishers, 1970.

Darlington, R.D. "Another Look at 'Cultural Fairness'." Journal of Educational Measurement, 1971, 8, 71-82.

Eells, K., Davis, A., Havighurst, R.J., Herrick, V.E., Tyler, R.W. Intelligence and Cultural Differences, Chicago: The University of Chicago Press, 1951.

Flaugher, R.L. "Some Points of Confusion in Discussing the Testing of Black Students." Research Memorandum, Educational Testing Service, RM - 73 - 5, Princeton, New Jersey, March 1973.

Flaugher, R.L. "The New Definitions of Test Fairness in Selection: Developments and Implications." Research Memorandum, Educational Testing Service, RM - 73 - 17, Princeton, New Jersey, September, 1973.

Franklin, A.J., et. al. "Ethnic Group Differences in Recall Based on the Composition of Task Content." (Unpublished manuscript), The Rockefeller University, Spring, 1974.

Goffman, E. Stigma. Englewood Cliffs, New Jersey: Prentice Hall, 1963.

Hertzig, M., Birch, H., and Thomas, A. "Class and Ethnic Differences in the Responsiveness of Preschool Children to Cognitive Demands." Monographs of the Society for Research in Child Development, 33 (1969), 1-69.

Jones, R. (ed.). Black Psychology, New York: Harper & Row Publishers, 1972.

Sarason, S.B. Anxiety in Elementary School Children, New York: John Wiley & Sons, 1960.

Thorndike, R.L. "Concepts of Culture Fairness." Journal of Educational Measurement, 1971, 8, 63-70.

Williams, R.L. "Abuses and Misuses in Testing Black Children." The Counseling Psychologist, 1971, 2 (1), 62-73.